

HYPER TEXT RETRIEVING DEVICE

Publication number: JP9311870 (A)

Publication date: 1997-12-02

Inventor(s): ISHIKAWA MIKITO; SATO MITSUHIRO; HOSHIDA MASANORI;
NOGUCHI YOSHIHIRO; YASUKAWA HIDEKI +

Applicant(s): MATSUSHITA ELECTRIC IND CO LTD +

Classification:

- **international:** **G06F12/00; G06F17/21; G06F17/30; G06F12/00; G06F17/21;**
G06F17/30; (IPC1-7): G06F12/00; G06F17/21; G06F17/27;
G06F17/30

- **European:** G06F17/30G4

Application number: JP19960149783 19960522

Priority number(s): JP19960149783 19960522

Also published as:

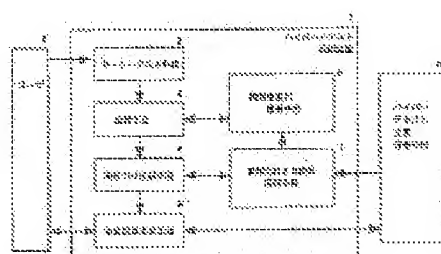
JP3108015 (B2)
EP0809197 (A2)
EP0809197 (A3)
EP0809197 (B1)
US5848407 (A)

[more >>](#)

Abstract of JP 9311870 (A)

PROBLEM TO BE SOLVED: To provide a hyper text retrieving device capable of sequencing the significance of hyper text documents most of which needs to include documents being not always independent of each other but in a referring relation to be consistent in terms of contents without erroneously evaluating, and allowing a user to smoothly select-designating in the sequenced result.

SOLUTION: A document table developing means 7 with a referring source is provided in the hyper text retrieving device 1 to develop a document table with a reference source making the identifier of the referring source document for previously analyzing the group of the hyper text documents and referring to the pertinent document for each document and an anchor sentence within the referring document into a tabular format.; Thereby, at the time of processing a sequence generation means 4, the referring relation between each document is known without inquiring to a hyper text document managing means 8 every time so that it is possible to evaluate-process significance integrating the documents in the referring relation and to display-process a sequencing result integrating the documents in the referring relation without delaying.



Data supplied from the **espacenet** database — Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-311870

(43) 公開日 平成9年(1997)12月2日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30			G 0 6 F 15/403	3 4 0 Z
12/00	5 1 5		12/00	5 1 5 M
	5 3 5			5 3 5 M
	5 4 7			5 4 7 H
17/27			15/20	5 5 0 F
審査請求 未請求 請求項の数12 F I (全 17 頁) 最終頁に続く				

(21) 出願番号 特願平8-149783

(22) 出願日 平成8年(1996)5月22日

(71) 出願人 000003821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 石川 幹人

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 佐藤 光弘

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 星田 昌紀

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74) 代理人 弁理士 役 昌明 (外2名)

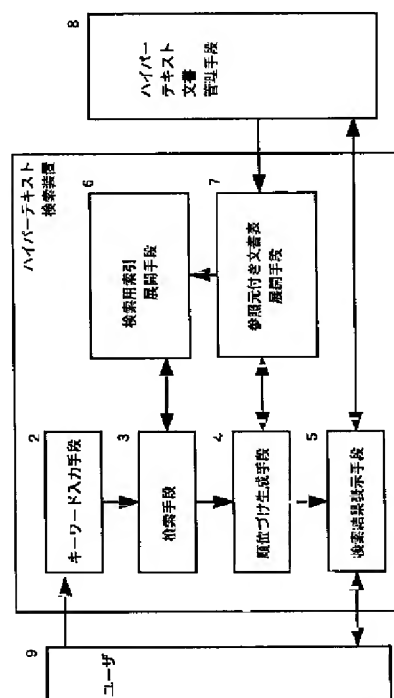
最終頁に続く

(54) 【発明の名称】 ハイパーテキスト検索装置

(57) 【要約】

【課題】 各文書が必ずしも独立でなく参照関係にある文書を含めて初めて一貫した内容となっていることが多いハイパーテキスト文書の重要度合を誤評価せずに順位づけができるようにすると共に、順位づけされた結果におけるユーザの選択指定が円滑に行なえるようにしたハイパーテキスト検索装置を提供する。

【解決手段】 ハイパーテキスト検索装置1内に参照元付き文書表展開手段7を設け、あらかじめハイパーテキスト文書群を解析し各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にした参照元付き文書表を展開しておくことにより、順位づけ生成手段4の処理に際して、各文書の参照関係がハイパーテキスト文書管理手段8にいちいち問い合わせることなく判明するため、参照関係にある文書を統合した重要度合の評価処理や、参照関係にある文書を統合した順位づけ結果の表示処理を遅滞なく行なえる。



【特許請求の範囲】

【請求項1】 ユーザの入力したキーワードに基づいてハイパーテキスト文書を検索し、その結果を順位づけして表示する装置において、参照元付き文書表展開手段を備え、前記参照元付き文書表展開手段にあらかじめ各文書の参照元を表形式に展開しておくことで、参照元文書の情報を加味した順位づけ表示を可能とするハイパーテキスト検索装置。

【請求項2】 ハイパーテキスト文書管理手段によって管理されるハイパーテキストによる参照構造を持った文書群を解析し、各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にして参照元付き文書表として保持する参照元付き文書表展開手段と、前記参照元付き文書表内の各文書の情報に基づき索引を作成して保持する検索用索引展開手段と、ユーザからの検索要求を受け付けるキーワード入力手段と、前記検索用索引を調べ当該キーワードを含む文書群を検索する検索手段と、前記検索手段によって検索された文書群に該当する前記参照元付き文書表内の文書情報に基づき当該文書群に含まれる各文書の重要度合を算出し順位づけて表示データとする順位づけ生成手段と、当該順位づけ表示データをユーザに表示する検索結果表示手段から構成されるハイパーテキスト検索装置。

【請求項3】 前記順位づけ生成手段が、順位づけされる各文書の見出しに添えて、当該文書の参照元文書の見出しを前記参照元付き文書表内から取り出して選択表示データとすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項4】 前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内のある文書と他の文書が近い参照関係にある場合、当該両文書を同一の順位にまとめた選択表示データを生成することを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項5】 前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書の重要度合に、当該文書の参照元のアンカー文の重要度合も加えて当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項6】 前記順位づけ生成手段が、順位づけされる各文書の見出しに添えて、前記参照元付き文書表内から当該文書の先頭部分およびキーワード周辺部分を取り出して選択表示データとすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項7】 前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書について当該文書に当該文書の参照元文書の内容を合わせて重要度合を算出し順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項8】 前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書について当該文書内の見出し部分の重要度合を他の部分よりも大きく評価して当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項9】 前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書の重要度合に、当該文書内で他の文書に対するアンカー文となっている部分の重要度合を減じて当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項10】 前記参照元付き文書表展開手段が前記参照元付き文書表を作成するにあたり、一文書内に参照ラベルが設定されている場合、当該文書を参照ラベルの位置で分割して複数の文書として再定義することを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項11】 ハイパーテキスト文書管理手段によって管理されるハイパーテキストによる参照構造を持った文書群を解析し、各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にして参照元付き文書表として保持する参照元付き文書表展開手段と、前記参照元付き文書表内の各文書の情報に基づき索引を作成して保持する検索用索引展開手段と、ユーザからの検索要求を受け付けるキーワード入力手段と、前記検索用索引を調べ当該キーワードを含む文書群を検索する検索手段と、前記検索手段によって検索された文書群に該当する前記参照元付き文書表内の文書情報に基づき当該文書群に含まれる各文書の重要度合を算出し順位づけて表示データとする順位づけ生成手段と、当該順位づけ表示データをユーザに表示する検索結果表示手段から構成され、前記順位づけ生成手段が、順位づけされた文書群の中に高い重要度合で現れる単語群を同定し、当該単語群を前記キーワード入力手段に戻すようにして、当該単語群から新たなキーワードを選択入力することで再検索が起動されるようにしたことを特徴とする請求項2記載のハイパーテキスト検索装置。

【請求項12】 前記順位づけ生成手段が、順位づけされた文書群の中の高い重要度合で現れる単語群を同定する際、キーワードの近傍に出現する単語の重要度合を他の部分よりも大きく評価して当該単語の重要度合とすることを特徴とする請求項11記載のハイパーテキスト検索装置。

【発明の詳細な説明】**【0001】**

【発明の属する技術分野】本発明はハイパーテキスト検索装置に関し 特に大量のハイパーテキスト文書の中から、ユーザの要求に合致しそうな文書を検索して提示するハイパーテキスト検索装置に関する。

【0002】

【従来の技術】従来の、ユーザの要求に合致しそうな文書を検索して提示する装置は、図11に示すような文書検索装置10であった。検索用索引展開手段6は、文書管理手段11に格納されている大量の文書をあらかじめ解析し、内蔵の辞書に登録されている単語群内の各単語について、当該単語がどの文書に何回現れているかを調べ、当該単語の全文書における偏り（idf値）を求め、当該単語が出現する各文書について規格化出現頻度（tf値）を掛け合わせる $tf \times idf$ 評価値を算出して検索用索引を展開する。

【0003】キーワード入力手段2は、ユーザ9から複数のキーワード入力を受け付け、検索手段3に供する。検索手段3は、当該入力キーワードと合致する単語群を前記検索用索引から探し、当該入力キーワードが出現する文書の識別子と、当該文書における当該キーワードの $tf \times idf$ 評価値との対を複数個得て、順位づけ生成手段4に供する。

【0004】順位づけ生成手段4は、 $tf \times idf$ 評価値の高い順に当該文書群を整列させ、検索結果表示手段5に供する。検索結果表示手段5は、前記整列順に文書の識別タイトルを列挙したうえでユーザの選択指定を受け付け、選択指定される都度、選択指定された文書のみを文書管理手段11から呼び出して表示する。

【0005】以上のような構成の文書検索装置10は、ユーザが要求に沿ったキーワードを入力することで、大量の文書からユーザの要求に近い文書をより近い順に提示させることができる。

【0006】なお、 $tf \times idf$ 評価値の具体的な算出方法は、いくつか知られているが、たとえば、当該単語を含む文書数の全文書に対する割合を対数値にしたものを1から引いた値をidf値とし、当該文書中の当該単語の出現回数を当該文書の大きさを割った値をtf値として両者を掛け合わせることで算出される。

【0007】 $tf \times idf$ 評価値の詳細および、それを応用した文書検索装置については、「Gerard Salton: Introduction to Modern Information Retrieval, McGraw-Hill, 1983」に記載されている。

【0008】

【発明が解決しようとする課題】従来構成の文書検索装置であっても、ハイパーテキスト文書の検索は不可能ではない。しかしながら、ハイパーテキスト文書は各文書が必ずしも独立でなく、参照関係にある文書を含めて初めて一貫した内容となっていることも多い。そうしたハイパーテキスト文書の検索を従来の文書検索装置で行なうと、ユーザにとっての文書の重要度合が誤って評価されて適切な順位づけができない問題や、順位づけされた結果におけるユーザの選択指定が円滑に行なえない問題が発生する。

【0009】特に、ハイパーテキストマークアップ言語（HTML）で書かれたインターネットのワールドワイドウ

ェブ（WWW）上の文書群では、一連の内容が参照関係にある複数の文書にわたっている傾向が顕著で、問題が大きい。

【0010】本発明は、こうした問題を解決し、ワールドワイドウェブ（WWW）上の文書群などの、これまで問題が大きかったハイパーテキスト文書群に対しても検索の利便性が高い文書検索装置を提供することを目的とする。

【0011】

【課題を解決するための手段】本発明のハイパーテキスト検索装置は、あらかじめハイパーテキスト文書群を解析して、各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にした参照元付き文書表を展開しておくことで、参照元の情報を加味した順位づけ表示を可能とするようにしたものである。

【0012】本発明によれば、検索および順位づけに際して、各文書がどの文書によって参照されているかがハイパーテキスト文書管理手段にいちいち問い合わせることなく判明するため、参照関係にある文書を統合した重要度合の評価処理や、参照関係にある文書を統合した順位づけ結果の表示処理を遅滞なく行なえる。その結果、ユーザの要求に合致したより適切な順位づけや、ユーザの選択指定が容易な順位づけ表示が実現できる。

【0013】

【発明の実施の形態】本発明の請求項1記載の発明は、ユーザの入力したキーワードに基づいてハイパーテキスト文書を検索し、その結果を順位づけして表示する装置において、参照元付き文書表展開手段を備え、前記参照元付き文書表展開手段にあらかじめ各文書の参照元文書を表形式に展開しておくことで、参照元の情報を加味した順位づけ表示を可能とするハイパーテキスト検索装置としたものであり、各文書がどの文書によって参照されているかがハイパーテキスト文書管理手段にいちいち問い合わせることなく、参照元付き文書表展開手段に展開された参照元文書の表を引くだけで判明し、参照関係にある文書群を統合した重要度合の評価処理や、参照関係にある文書群を統合した順位づけ結果の表示データ生成を高速に行なえるという作用を有する。

【0014】本発明の請求項2記載の発明は、ハイパーテキスト文書管理手段によって管理されるハイパーテキストによる参照構造を持った文書群を解析し、各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にして参照元付き文書表として保持する参照元付き文書表展開手段と、前記参照元付き文書表内の各文書の情報に基づき索引を作成して保持する検索用索引展開手段と、ユーザからの検索要求を受け付けるキーワード入力手段と、前記検索用索引を調べ当該キーワードを含む文書群を検索する検索手段と、前記検索手段によって検索された文書群に該当する

前記参照元付き文書表内の文書情報に基づき当該文書群に含まれる各文書の重要度を算出し順位づけて表示データとする順位づけ生成手段と、当該順位づけ表示データをユーザに表示する検索結果表示手段から構成されるハイパーテキスト検索装置としたものであり、各文書がどの文書によって参照されているかがハイパーテキスト文書管理手段にいちいち問い合わせることなく、参照元付き文書表展開手段に展開された参照元文書の表を引くだけで判明し、参照関係にある文書群を統合した重要度の評価処理や、参照関係にある文書群を統合した順位づけ結果の表示データ生成を高速に行なえるようになるとともに、ユーザが要求に沿ったキーワードを入力することで、大量のハイパーテキスト文書からユーザの要求に近いハイパーテキスト文書群がより近い順に高速に提示されるという作用を有する。

【0015】本発明の請求項3記載の発明は、前記順位づけ生成手段が、順位づけされる各文書の見出しに添えて、当該文書の参照元文書の見出しを前記参照元付き文書表内から取り出して選択表示データとすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、参照関係になっている文書群は通常深い関連性がある内容を持っているので、検索された文書に関連する文書のユーザによる選択呼び出しが容易となる作用を有する。

【0016】本発明の請求項4記載の発明は、前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内のある文書と他の文書が近い参照関係にある場合、当該両文書を同一の順位にまとめた選択表示データを生成することを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、参照関係になっている文書群は通常深い関連性がある内容を持っているので、検索された文書群とそれらの関連文書群の一覧性を高め、同時にユーザの同一文書の重複呼び出しを防ぐという作用を有する。

【0017】本発明の請求項5記載の発明は、前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書の重要度合に、当該文書の参照元のアンカー文の重要度合も加えて当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、アンカー文は通常参照先の文書の特徴を良く記述しているので、検索された文書群の順位づけの信頼性を高めるといふ作用を有する。

【0018】本発明の請求項6記載の発明は、前記順位づけ生成手段が、順位づけされる各文書の見出しに添えて、前記参照元付き文書表内から当該文書の先頭部分およびキーワード周辺部分を取り出して選択表示データとすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、ユーザが順位づけされた各文書の概要を当該文書を選択呼び出しすることなしに理

解でき、ユーザによる検索結果の取捨選択が容易となるという作用を有する。

【0019】本発明の請求項7記載の発明は、前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書について当該文書に当該文書の参照元文書の内容を合わせて重要度を算出し順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、ユーザの検索要求にかかわる内容が参照関係にある複数文書に分離している場合に、当該複数文書が分離して扱われて順位づけの低下を招くことを防ぐという作用を有する。

【0020】本発明の請求項8記載の発明は、前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書について当該文書内の見出し部分の重要度合を他の部分よりも大きく評価して当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、見出し部分が文書の特徴を通常良く記述しているので、検索された文書群の順位づけの信頼性を高めるといふ作用を有する。

【0021】本発明の請求項9記載の発明は、前記順位づけ生成手段が、文書群の順位づけに際し前記参照元付き文書表を調べ、当該文書群内の各文書の重要度合に、当該文書内で他の文書に対するアンカー文となっている部分の重要度合を減じて当該文書の重要度合として順位づけすることを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、参照先ばかりが列挙されておりユーザが要求する内容が乏しい文書の順位づけを低く抑えるという作用を有する。

【0022】本発明の請求項10記載の発明は、前記参照元付き文書表展開手段が前記参照元付き文書表を作成するにあたり、一文書内に参照ラベルが設定されている場合、当該文書を参照ラベルの位置で分割して複数の文書として再定義することを特徴とする請求項2記載のハイパーテキスト検索装置としたものであり、内部に複数の意味まとまりを持つ長大な文書を各意味まとまりごとに検索し順位づけを行なえるという作用を有する。

【0023】本発明の請求項11記載の発明は、ハイパーテキスト文書管理手段によって管理されるハイパーテキストによる参照構造を持った文書群を解析し、各文書ごとに当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にして参照元付き文書表として保持する参照元付き文書表展開手段と、前記参照元付き文書表内の各文書の情報に基づき索引を作成して保持する検索用索引展開手段と、ユーザからの検索要求を受け付けるキーワード入力手段と、前記検索用索引を調べ当該キーワードを含む文書群を検索する検索手段と、前記検索手段によって検索された文書群に該当する前記参照元付き文書表内の文書情報に基づき当該文書群に含まれる各文書の重要度合を算出し順位づけて表示デ

ータとする順位づけ生成手段と、当該順位づけ表示データをユーザに表示する検索結果表示手段から構成され、前記順位づけ生成手段が、順位づけされた文書群の中に高い重要度合で現れる単語群を同定し、当該単語群を前記キーワード入力手段に戻すようにして、当該単語群から新たなキーワードを選択入力することで再検索が起動されるようにしたことを特徴とするハイパーテキスト検索装置としたものであり、検索結果を見ながらの絞り込みを狙った再検索や興味の湧いたキーワードによる新たな検索を、キーボードを使うことなしにポインティングデバイスのみを用いたキーワード入力が可能であるという作用を有する。

【0024】本発明の請求項12記載の発明は、前記順位づけ生成手段が、順位づけされた文書群の中に高い重要度合で現れる単語群を同定する際、当該入力キーワードの当該文書における出現位置の近傍に現れる単語の重要度合を他の部分に出現する単語よりも大きく評価して当該単語の重要度合とすることを特徴とする請求項11記載のハイパーテキスト検索装置としたものであり、キーワードの近傍に出現する単語が通常そのキーワードと関連性が高いので、関連キーワードとしての利用価値が高い単語群が選定列挙される可能性を高めるという作用を有する。

【0025】以下、本発明の実施の形態について、例として、インターネットのワールドワイドウェブ（WWW）上に分散管理されている文書群から、特定の文書群をキーワード検索するハイパーテキスト検索装置について図1から図10を用いて説明する。

【0026】一般にハイパーテキスト文書は図2のように、複数の参照関係にある文書80～86によって構成されている。文書80～84にはほとんどころにアンカー文800～806があり参照先の文書の識別子あるいは文書格納位置が埋め込まれている。ユーザはいわゆるブラウザと呼ばれる文書閲覧選択手段を用い、例えば文書81を閲覧しながらアンカー文801の位置をいわゆるポインティングデバイスで選択することにより、参照先の文書83を呼び出し閲覧するといった方法でハイパーテキスト文書を利用する。文書81に対して文書83を参照先文書、逆に文書83に対して文書81を参照元文書と呼ぶ。

【0027】ワールドワイドウェブWWW上のハイパーテキスト文書群はハイパーテキストマークアップ言語（HTML）で書かれており、ひとつの文書はページと呼ばれて、単に文字列だけでなく、画像やプログラムをも含んでいる。HTMLでは、nouka.htmlというファイルに格納された文書81内の「りんご生産農家」という文字列に対して、ringo.htmlというファイルに格納された文書83への参照を埋め込んでアンカー文とするには、文書81内の該当文字列部分を「`りんご生産農家`」のように表記すればよい。その場合、参照先文書83には何も記さなくて良いので、文書83が作成され

たあとに、文書83が格納されている計算機からはネットワーク上で遠く離れた位置にある計算機内に文書82が作成され、文書82が文書83の参照元文書のひとつになっているという場合も起こりうる。

【0028】（第1の実施の形態）図1は、本発明のハイパーテキスト検索装置1を含む全体構成を示すブロック図である。図1のハイパーテキスト文書管理手段8は、ワールドワイドウェブWWWの場合ではネットワーク上に広域に分散された多数の計算機によって分散管理されている。そのため、本発明のハイパーテキスト検索装置1の参照元付き文書表展開手段7は、関連文書収集機能（一般的にウェブロボットと呼ばれる）を装備しており、幾つかの文書格納位置アドレス（ユニバーサルリソースロケータ、URL）を与えると当該URLが指し示す文書から順にアンカー文に埋め込まれているURLを手がかりに参照先の文書を次々と収集する。そして収集された文書に、収集順などに基づく重複しない文書識別子を付す。なお、画像やプログラムは除いて文字列のみを収集すると収集時間を節約でき、また、すでに収集した文書のURLをリストにしておき、当該リストにあるURLが指し示す文書は重ねて収集しないこととすると、図2に示す文書84から文書83を参照するアンカー文806のような戻りの参照関係があるときに重複して文書を収集してしまうことが避けられる。

【0029】次に図1の参照元付き文書表展開手段7は、収集した文書群から図3に示すような参照元付き文書表を、次の手順で作成する。収集した文書数に相当する個数の文書情報記入欄70～72を作成し、当該各文書情報記入欄に文書識別子の番号と文書格納位置のURLアドレスを記入する。収集した各文書について、当該HTML文書内の文字列を調べ、「`<title>うちのりんご</title>`」などと記されている部分から「うちのりんご」などのタイトルを抽出して記入したうえ、BODY部に存在する文字列から「`<`」と「`>`」で囲まれた部分であるHTMLタグの文字列を除いてテキスト本文として記入する。収集した各文書について、当該文書内の文字列を調べアンカー文があったならば、当該アンカー文の参照先文書に該当する文書情報記入欄の参照元リストに、当該アンカー文をもつ文書の識別子と当該アンカー文とを対にして記入する。

【0030】このように参照元付き文書表展開手段7は、ハイパーテキスト文書管理手段8に格納されている大量の文書をあらかじめ収集してそれを整理し、各文書ごとに当該文書の識別子、文書格納位置、タイトル、テキスト本文などに加えて、当該文書を参照する参照元文書の識別子および参照元文書内のアンカー文を表形式にして参照元付き文書表として展開し、保持するものである。

【0031】なお、以上では参照元付き文書表の作成を全ての文書収集が終わった後に行なうとして説明したが、

前記文書収集時にアンカー文を調べるのに合わせて、参照元付き文書表の作成を順次行なう方法をとると効率化が図れる。

【0032】また本発明の実施の形態におけるtf値は、単語の文書内出現回数そのものを使用し、文書の大きさを規格化しないとしている。これに対し文書の大きさを規格化するtf値を使う場合は、各文書情報記入欄に文書の大きさも算出して展開しておく必要がある。

【0033】図1の検索用索引展開手段6は、数万語の辞書を有しており、図4に示すような検索用索引を、次の手順で作成する。辞書に登録されている単語数に相当する個数の単語情報記入欄60～62を作成し、当該単語情報記入欄それぞれに単語を記入する。前記参照元付き文書表内の全ての文書情報記入欄の内容を単語切りし、各単語に相当する文字列が存在する文書の識別子と出現位置（出現場所）を、該当する単語の単語情報記入欄に記入する。

【0034】図4は、図3の文書83の内容を単語切りした後の、「りんご」という単語の単語情報記入欄を示している。（TITLE,4）はタイトルの4文字目に「りんご」があること、（BODY,1,33,43）はテキスト本文の1文字目、33文字目、43文字目に「りんご」があること、（000082,5）は、文書82のアンカー文の5文字目に「りんご」があることを意味する。また各単語の出現文書数（あるいはidf値）や各文書内出現回数（あるいはtf値）などのあらかじめ算出しておける数値は、検索用索引展開時に算出して当該単語情報記入欄に記入しておくこと、検索時の処理時間が節約できる。さらに各単語情報記入欄に登録されている文書については、当該文書の文書情報記入欄の単語リストに図3の最下段にあるように当該単語を並べておくと、文書情報から容易に単語情報を引くことができる。

【0035】このように検索用索引展開手段6は、前記参照元付き文書表に格納されているテキスト本文、タイトル、アンカー文などの文字列をあらかじめ解析し、内蔵の辞書に登録されている単語群内の各単語について、当該単語がどの文書内の文字列のどの位置に現れているかを記した検索用索引を展開し、保持するものである。

【0036】図1のキーワード入力手段2は、テキストボックス機能と当該テキストボックスの内容を返す検索開始用ボタンを備えて「検索のページ」などと題したHTML文書を用いて実現される。ユーザは、自分のクライアント計算機で起動したモザイクあるいはネットスケープなどのWWWブラウザで当該HTML文書呼び出し、テキストボックスに単一のキーワードを入力するかあるいは複数のキーワードをスペースで区切って列挙して入力し、続けて検索開始用ボタンを押すことにより、キーワードの入力がなされる。

【0037】このようにキーワード入力手段2は、ユーザ9から複数のキーワード入力を受け付け、検索手段3

に供給するものである。なお、キーワードの候補となる複数の単語をボタンとして列挙し選択入力可能としておくと、キーボードから任意文字列を入力するのが不得手のユーザにも、当該キーワード候補の単語範囲で、キーボードを使わずともポインティングデバイスを用いたキーワード入力が可能である。

【0038】図1の検索手段3は、前記キーワード入力手段2によって入力されたキーワード群と同一単語の単語情報記入欄を前記検索用索引から取り出す。

【0039】このように検索手段3は、当該入力キーワードと合致する単語群を前記検索用索引から探し、当該入力キーワードが出現する文書の識別子と、当該キーワードが出現する当該文書の文字列内の位置情報との組を複数個得て、順位づけ生成手段4に供するものである。

【0040】図1の順位づけ生成手段4は、当該キーワードが出現する各文書についてtf×idf評価値を算出して当該文書の重要度合とし、当該重要度合が高い順に当該文書群を順位づける。入力されたキーワードが複数ある場合は、各々のキーワードに関して出現する文書のtf×idf評価値を算出して重要度合とし、全文書にわたり重要度合を比較して順位づける。

【0041】なお、複数種のキーワードを含む文書について各キーワードのtf×idf評価値を足し合わせて当該文書の重要度合として順位づける。また、複数種のキーワードを含む文書についてtf×idf評価値を足し合わせる際に、キーワードを2種含む時はさらに2倍するとか、キーワードを3種含む時はさらに3倍するとかの加点（重み付け）を行なうとキーワードの共起による順位づけが上がり、ユーザの要求に沿うことが多くなる。さらに2種のキーワードが文書内で20文字より近い距離にある場合はさらに2倍するなどの、キーワードの文書内近接共起による加点（重み付け）を行なうと、キーワード共起による順位づけがさらに強く効き、ユーザの要求に沿うことが多くなる。

【0042】順位づけ生成手段4は、さらに、順位づけされた順で各文書の見出しを列挙したHTML文書を表示用に作成し、検索結果表示手段5に供する。見出しとは、各文書のタイトルあるいはその参照元文書のアンカー文の文字列であり、個々の見出しには当該文書のURLを埋め込んでアンカー文としておく。

【0043】このように順位づけ生成手段4は、前記参照元付き文書表から参照関係にある文書群を導き、当該文書群を統合した重要度合の評価処理および順位づけ並びに順位づけ結果の表示データ生成を行ない、検索結果表示手段5に供するものである。

【0044】図1の検索結果表示手段5は、ユーザのクライアント計算機で起動したモザイクあるいはネットスケープなどのWWWブラウザで実現され、表示用に作成されたHTML文書を受け付けクライアント計算機に画面表示すると同時に、ポインティングデバイスで見出しを選択

指定することで、見出しに対応する実体の文書が当該見出しに埋め込まれているURLに基づいて呼び出し可能となる。

【0045】このように検索結果表示手段5は、前記表示データを画面表示したうえでユーザの選択指定を受け付け、選択指定される都度、選択指定された文書のみをハイパーテキスト文書管理手段8から呼び出して表示するものである。

【0046】以上のように本発明の第1の実施の形態によれば、各文書がどの文書によって参照されているかがハイパーテキスト文書管理手段8にいちいち問い合わせることなく、参照元付き文書表展開手段7に展開された参照元文書の表を引くだけで判明し、参照関係にある文書群を統合した重要度合の評価処理や、参照関係にある文書群を統合した順位づけ結果の表示データ生成を高速に行なえるようになるとともに、ユーザが要求に沿ったキーワードを入力することで、大量のハイパーテキスト文書からユーザの要求に近いハイパーテキスト文書群がより近い順に高速に提示されるという有利な効果を奏する。

【0047】(第2の実施の形態)本発明の第2の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。たとえば図2に示される参照関係にある文書群から文書83、文書85、文書86が検索され、各々4位、18位、19位に順位づけられた場合を考える。

【0048】順位づけ生成手段4は、順位づけ結果生成時に、検索された各文書(図6の星印を付した文書)の参照元文書を参照元付き文書表を用いてたどっていき、たとえば1つ上と2つ上の参照元文書の見出しを添えた順位づけ結果表示が、図6のように生成できる。文書83の参照元関係にある文書は文書81と文書82の複数があるが、文書格納位置の部分文字列同士をマッチさせることで、より文書格納位置に近い文書を判定して参照元文書をひとつに絞っている。このように参照元文書の見出しを添えておけば、図6の結果表示から、19位にある文書86のようにアンカー文をひとつも持たない文書(図2参照)を選択し呼び出してしまっても、再び図6の結果表示に戻れば、深い関連性がある内容をもつであろう参照元の文書83や文書81が容易に選択呼び出し可能となる。

【0049】以上のように本発明の第2の実施の形態によれば、順位づけ生成手段4において、前記参照元付き文書表から検索された文書の参照元関係にある文書群の見出しを取り出し、順位づけ結果の表示データ生成の際に、検索された文書の見出しだけでなく当該参照元文書の見出しも合わせて表示データとし、検索結果表示手段5に供することにより、参照関係になっている文書群は通常深い関連性がある内容を持っているので、検索された文書に関連する文書のユーザによる選択呼び出しが容

易となる有利な効果を奏する。

【0050】(第3の実施の形態)本発明の第3の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。また前記第2の実施の形態で行なった図6のような表示では問題が多少残ることに対して考慮したものである。たとえば図2および図6において、4位にある文書83を呼び出して読み、さらに文書83にあるアンカー文から文書85を呼び出して読んだとする。その後、5位、6位と読み進むと、すでに読み終わっている文書85が18位に出現するので読み終わっていると気づかずに再度呼び出してしまったり、19位で文書86を呼び出して読んでも、文書83を読んだから時間がたっているので文脈がよく理解できなかったりする可能性がある。

【0051】そこで、本発明の第3の実施の形態では、図1の順位づけ生成手段4において、前記参照元付き文書表から検索された文書の参照元関係にある文書群の識別子を取り出し他の検索された文書の識別子と同一性を判定することで、検索された文書同士が近い参照関係にあることを認定し、順位づけ結果の表示データ生成の際に、近い参照関係が認定された当該文書群の見出しを同一の順位の表示データにして検索結果表示手段5に供するようにしている。その結果、図6の4位の文書群と18位の文書群と19位の文書群のように共通の文書を含む文書群は、それらの文書識別子から判定して、図8のように4位に一体化して表示するようにすると、検索された文書83、文書85、文書86が参照関係の文脈を保持しながら続いて読むことができるので、ユーザによる検索結果の一覧性を高めると同時に、ユーザの同一文書の重複呼び出しを防ぐことができる。

【0052】なお、以上では一体化表示時に、一体化される文書のうち最も重要度合の大きな文書の重要度合を、一体化された文書群の重要度合として順位づけるとして説明したが、一体化された文書群の重要度合を、一体化された文書群の重要度合の和として順位づけしてもよい。

【0053】(第4の実施の形態)本発明の第4の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。すなわち、本発明の第4の実施の形態では、たとえば入力キーワード「りんご」に関する検索では、図4の「りんご」の単語情報記入欄が指し示している「りんご」の出現位置が文書81および文書82のアンカー文であっても、当該出現回数2回を文書83のテキストに出現した「りんご」の出現回数4回に加えて6回としidf×tf評価値を計算して順位づけると、アンカー文では通常参照先の文書の特徴を記述したキーワード(この場合は「りんご」)が何度も現れるので、検索された文書83の順位づけが高くな

り、「りんご」に関する検索の信頼性を高められる。

【0054】つまり、第4の実施の形態では、図1の順位づけ生成手段4において、検索手段3から供された入力キーワードの出現位置が前記参照元付き文書表内の参照元文書のアンカー文であっても、当該出現回数を参照先の文書内のテキストに出現した当該入力キーワードの出現回数に加えて当該参照先文書の重要度合を計算して順位づけたうえで表示データを生成し、検索結果表示手段5に供するようにする。これにより、アンカー文は通常参照先の文書の特徴を良く記述しているので、検索された文書群の順位づけの信頼性を高める有利な効果を奏する。

【0055】（第5の実施の形態）本発明の第5の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。すなわち、本発明の第5の実施の形態では、順位づけ結果生成時に、単語情報記入欄から得られる入力キーワードのテキスト本文内の出現位置にもとづき、当該入力キーワードを含む文あるいはフレーズを抜き出して当該文書のテキスト本文の先頭文あるいは先頭フレーズに続けて表記した要約文を形成し、図7に示すように、各文書の見出しに添えて表示データとすると、ユーザが順位づけされた文書をいちいち選択呼び出しすることなしに、順位づけ結果表示画面に表記された各文書の要約文を読むだけで当該文書の概要を理解でき、ユーザによる検索結果の取捨選択が容易となる。

【0056】つまり、第5の実施の形態では、図1の順位づけ生成手段4において表示データを生成するに際し、検索手段3から供された入力キーワードの前記参照元付き文書表内の出現位置にもとづき、当該入力キーワードを含む文あるいはフレーズを抜きだして当該文書のテキスト本文の先頭文あるいは先頭フレーズに添えて表示データとし、検索結果表示手段5に供するようにする。これにより、ユーザが順位づけされた各文書の概要を当該文書を選択呼び出しすることなしに理解でき、ユーザによる検索結果の取捨選択が容易となる有利な効果を奏する。

【0057】なお、以上ではひとつの文書にキーワードが何度現れても当該入力キーワードを含む文あるいはフレーズを全て抜きだして要約文を生成するとして説明したが、5種までの複数キーワード検索は1種当たり3箇所までしか抜き出さないとか、10種までの複数キーワード検索は1種当たり2箇所までしか抜き出さないとか、10種を越える複数キーワード検索は1種当たり1箇所までしか抜き出さないとかの制限を設けると、要約文が長大になって順位づけの一覧性が低下するのを防ぐことができる。

【0058】（第6の実施の形態）本発明の第6の実施の形態は、前記第1の実施の形態と基本的な構成は同じ

である。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。すなわち、ワールドワイドウェブWWWの文書（ページ）では、各文書が細切れになっており、一文書内に十分な文脈が記述されていないことが多い。たとえば文書83の「りんご」が「青森りんご」であるにもかかわらず、「青森」の記述は文書83にはなく参照元の文書81にあることがある。

【0059】本発明の第6の実施の形態では、たとえば入力キーワード「りんご」と「青森」に関する検索において文書83の重要度合を算出するにあたり、文書83と参照元文書81を合わせた文書の重要度合と、文書83と参照元文書82を合わせた文書の重要度合とを比較し、「青森」という単語が入っていて重要度合が大きい文書81と文書83を合わせた文書の重要度合を文書83の重要度合とすると、文書83の順位づけの低下を防ぐことができる。

【0060】つまり、第6の実施の形態では、図1の順位づけ生成手段4において、検索手段3によって検索された文書群内の各文書について前記参照元付き文書表から当該文書の参照元関係にある文書群を特定し、当該検索された文書の内容と当該参照元関係にある文書の内容を複合した内容で重要度合を算出し、当該参照元関係にある文書群のうち当該検索された文書と複合して最も評価が高くなる文書を選び、当該検索された文書の重要度合に代えて当該複合文書の重要度合を用いて順位づけたうえで表示データとし、検索結果表示手段5に供するようにする。これにより、ユーザの検索要求にかかわる内容が参照関係にある複数文書に分離している場合に、当該複数文書が分離して扱われて順位づけの低下を招くことを防ぐという有利な効果を奏する。

【0061】（第7の実施の形態）本発明の第7の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。すなわち、本発明の第7の実施の形態では、入力キーワード「りんご」の出現位置が文書のタイトルあるいは参照元文書のアンカー文などの見出し部分の文字列にあるとき、文書83内のテキスト全体に出現した当該入力キーワードの出現回数6回のうちアンカー文とタイトルに出現した回数3回を2倍にし、出現回数を9回として当該文書の重要度合を計算するなどすると、見出し部分が文書の特徴を通常良く記述しているので、文書83の順位づけの信頼性をさらに高められる。なお、見出し部分にはアンカー文とタイトルに加え、HTML文では「<h1>」と「</h1>」で囲まれる文字列に代表される小見出しを含めても良い。

【0062】つまり、本発明の第7の実施の形態では、図1の順位づけ生成手段4において、検索手段3から供された入力キーワードの出現位置が前記参照元付き文書表内の見出し部分の文字列にあるとき、当該出現回数を数倍したうえで文書内のテキスト全体に出現した当該入

力キーワードの出現回数に加えて当該文書の重要度合を算出し、当該文書を順位づけたいうでで表示データを生成し、検索結果表示手段5に供するようにする。これにより、見出し部分が文書の特徴を通常良く記述しているの、検索された文書群の順位づけの信頼性を高める有利な効果を奏する。

【0063】(第8の実施の形態)本発明の第8の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の順位づけ生成手段4の機能を若干異にしている。すなわち、ワールドワイドウェブWWWの文書(ページ)には、他の文書を参照するアンカー文が多数列挙されているいわゆるリンクページが存在する。リンクページは検索表示されても検索意図に合った有益な情報がないことが多い。

【0064】本発明の第8の実施の形態では、たとえば文書82のテキスト本文が「りんご」に関するアンカー文10個のみのリンクページであり「りんご」の出現回数がテキスト本文で10回であった場合でも、「りんご」に関する検索において文書83での「りんご」の出現回数を調べるにあたり、参照元文書のアンカー文に「りんご」が見い出される都度、文書82のテキスト本文での「りんご」の出現回数を1回分減じることを繰り返すことにより、文書82のテキスト本文での「りんご」の出現回数を0回にまで低下させられ、リンクページの順位づけを低く抑えることができる。

【0065】つまり、本発明の第8の実施の形態では、図1の順位づけ生成手段4において、検索手段3から供された入力キーワードの出現位置が前記参照元付き文書表内の参照元文書のアンカー文であるとき、当該出現回数を当該参照元の文書内のテキスト本文に出現した当該入力キーワードの出現回数から減じて当該参照元文書の重要度合を計算して順位づけたいうでで表示データを生成し、検索結果表示手段5に供するようにする。こうすることにより、参照先ばかりが列挙されておりユーザが要求する内容が乏しい文書の順位づけを低く抑えることができる。

【0066】(第9の実施の形態)本発明の第9の実施の形態は、前記第1の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置1の参照元付き文書表展開手段7の機能を若干異にしている。すなわち、本発明の第9の実施の形態では、参照元付き文書表として展開するに際し、図5の文書87のように「」などのラベルがあり、かつ当該文書内からあるいは他の文書から「」などと参照されている場合には、当該文書を参照ラベルの位置の直前で分割し、ラベル以降を各々別の文書(文書88、文書89)として扱って参照元付き文書表を作っておくと、内部に複数の意味まとまりを持つ長大な文書を各意味まとまりごとに検索し順位づけを行なえる。なお、上に示されるような参照ラベルに加えて、「<h1>」と「</h1>」

で囲まれる文字列に代表される小見出しの位置の直前でもって分割し、各々別の文書として扱っても良い。

【0067】つまり、本発明の第9の実施の形態では、図1の参照元付き文書表展開手段7において、ハイパーテキスト文書管理手段8に格納されている大量の文書をあらかじめ収集整理し参照元付き文書表として展開するに際し、各文書に当該文書内からあるいは他の文書からの参照ラベルが付されているか判定し、参照ラベルが付されている場合には当該文書を参照ラベルの位置で分割し、各々別の文書として扱って参照元付き文書表を作っておき、前記検索用索引展開およびキーワード入力による検索と順位づけを行なうようにする。こうすることにより、内部に複数の意味まとまりを持つ長大な文書を各意味まとまりごとに検索し順位づけを行なえるという有利な効果を奏する。

【0068】(第10の実施の形態)図9は、図1の本発明のハイパーテキスト検索装置1と若干異なるハイパーテキスト検索装置91を含む全体構成を示すブロック図である。図9の構成と図1の構成とで異なる構成は、図9の順位づけ生成手段94の順位づけ結果がキーワード入力手段92に戻されている点である。その他の構成は、図1と同じなので構成の説明は省略することにする。

【0069】本発明の第10の実施の形態では、順位づけ結果生成時に、順位づけされた文書群内のたとえば上位10文書について、各文書の文書情報記入欄の単語リストを調べ、そこに存在する各単語について当該10文書内のtf×idf評価値を算出して各単語の重要度合とし、重要度合の高い順にたとえば10単語を選んでキーワード入力手段92に戻して、当該10単語を選択ボタンとしたHTML文書を生成する。そして図10に示すように、検索結果表示手段5は、WWWブラウザのネットスケープのフレーム機能を用いて表示画面を2分割し、画面上段に順位づけ結果表示窓50を画面下段に関連キーワード入力窓20を設け当該HTML文書を表示することにより、たとえば「りんご」の検索結果を見ながら、より検索意図に近い「特産品」で絞り込みを狙った再検索を起動したり、「ねぶた」などの興味の湧いた他のキーワードによる新たな検索を起動するのが、キーボードを使うことなしにポインティングデバイスのみを用いた入力で可能である。

【0070】なお、図10における「特産品」「ねぶた」などの単語ボタンは当該単語をテキストボックスに追加するJAVAスクリプトを動作させ、「クリア」のボタンはテキストボックスの単語群をクリアするJAVAスクリプトを動作させ、「初期状態」のボタンはテキストボックスの単語群を前回の検索に使用したキーワード群(この場合は「りんご」)に戻すJAVAスクリプトを動作させ、「再検索」ボタンでテキストボックスの単語群をキーワードとして検索が再起動されるように実装される。

【0071】なお、以上の説明における順位づけされた

文書群内の上位10文書とは、検索された文書（図8または図10の星印の文書）を想定したが、検索された文書の参照元文書（図8または図10の順位づけ表示に現れる文書群）までに広げて上位10位までに現れる全文書すると、参照関係にある文書から広く関連キーワードが収集できる。

【0072】つまり、第10の実施の形態では、図9の順位づけ生成手段94において、前記参照元付き文書表から順位づけされた文書群に存在する単語を調べ、各単語についての重要度合を前記検索用索引の情報から算出し、重要度合の高い順に数単語をキーワード入力手段92に戻して選択ボタンにした表示を、検索結果表示手段5の表示と同一画面内に配置するようにする。こうすることにより、検索結果を見ながらの絞り込みを狙った再検索や興味の湧いたキーワードによる新たな検索を、キーボードを使うことなしにポインティングデバイスのみを用いたキーワード入力が可能であるという有利な効果を奏する。

【0073】（第11の実施の形態）本発明の第11の実施の形態は、前記第10の実施の形態と基本的な構成は同じである。しかし、ハイパーテキスト検索装置91の順位づけ生成手段94の機能を若干異にしている。すなわち、本発明の第11の実施の形態では、関連キーワードの重要度合を算出するにあたり、各単語について前記検索用索引の情報から文書中の出現位置を調べ、当該入力キーワードの前後20文字などの近傍位置に出現した単語は、その出現回数を2倍にしたうえで $t \times i \times d \times f$ 評価値を求めて重要度合とし、その高い順に10単語を選定して関連キーワード入力窓20に順に表示すると、図10に示すように、「りんご」という入力キーワードの近傍に出現しやすく「りんご」に高い関連性を持つ「信州」「農家」などの単語が、入力キーワードの近傍には出現しにくいが関連文書にはよく出現する「長野」「オリンピック」などの単語よりも先に表示される。これにより、絞り込みを狙った再検索に使う単語が最初の方にあり、新たな検索の興味をひく単語が後の方にある傾向をもった単語群の列举表示が可能となり、当該単語群の関連キーワードとしての利用価値が高まる。

【0074】つまり、第11の実施の形態では、図9の順位づけ生成手段94において、前記参照元付き文書表から順位づけされた文書群に存在する単語群を調べ、各単語についての重要度合を前記検索用索引の情報から算出する際、当該キーワードの近傍に出現した単語の重要度合を数倍したうえで、重要度合の高い順に数単語を選定し列举するようにする。こうすることにより、キーワ

ードの近傍に出現する単語が通常そのキーワードと関連性が高いので、関連キーワードとしての利用価値が高い単語群が選定列举される可能性を高めることができる。

【0075】

【発明の効果】以上のように本発明によれば、参照関係にある文書を統合した重要度合の評価処理や、参照関係にある文書を統合した順位づけ結果の表示処理を効率良く行なえ、ユーザの要求に合致したより適切な順位づけや、ユーザの選択指定が容易な順位づけ表示が実現できるという有利な効果が得られる。

【図面の簡単な説明】

【図1】本発明の全体構成を示すブロック図、

【図2】ハイパーテキスト文書の構造を示す図、

【図3】収集した文書群から作成された参照元付き文書表を示す図、

【図4】参照元付き文書表内の全ての文書情報記入欄の内容を単語切りし、各単語毎に作成される検索用索引の一例を示す図、

【図5】ラベル付き文書の分割の様子を示す図、

【図6】順位づけ結果を表示する第1の図、

【図7】順位づけ結果を表示する第2の図、

【図8】順位づけ結果を表示する第3の図、

【図9】関連キーワードによる再検索機能を実現する構成を示すブロック図、

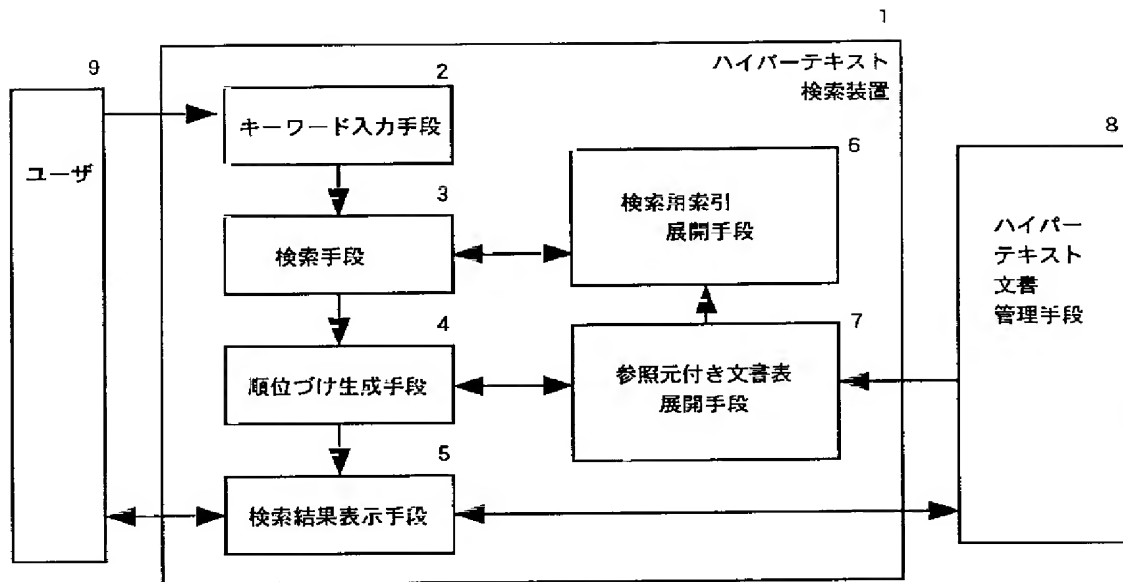
【図10】関連キーワード付き順位づけ結果を表示する図、

【図11】従来例の全体構成を示すブロック図である。

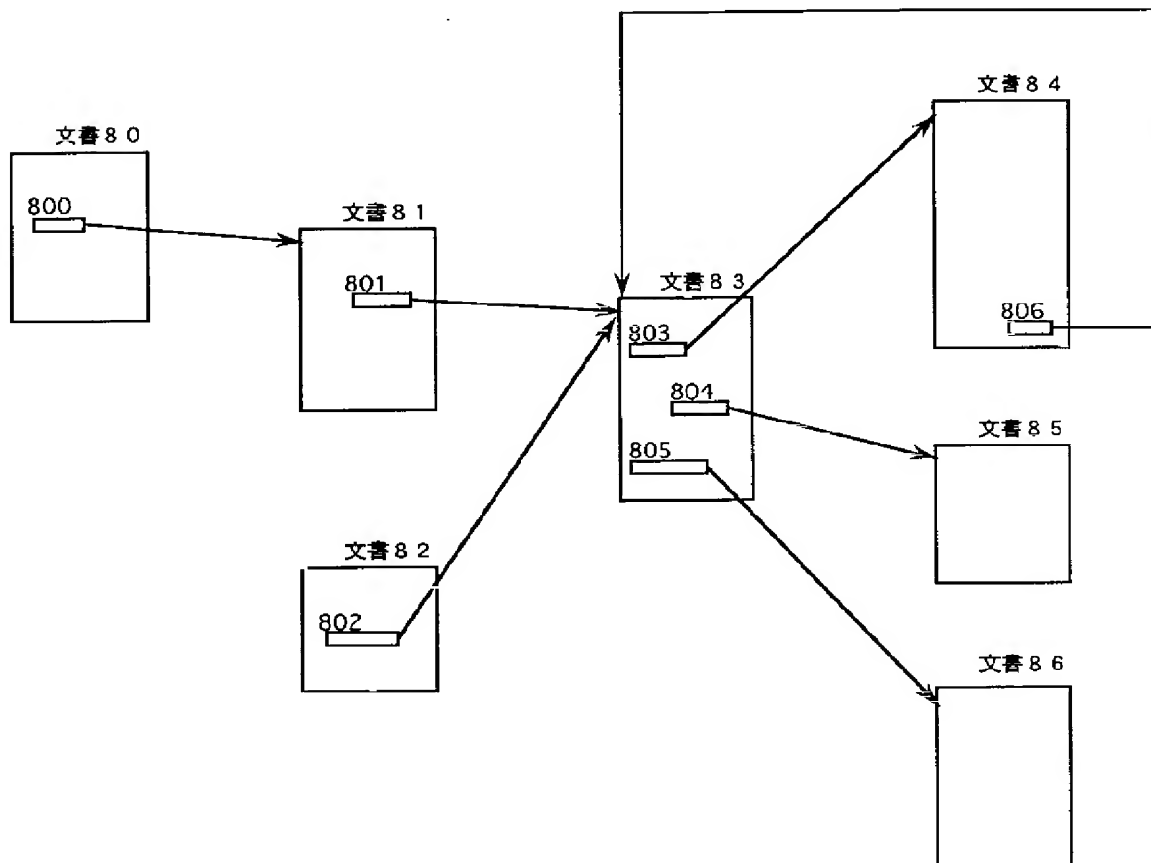
【符号の説明】

- 1、91 ハイパーテキスト検索装置
- 2、92 キーワード入力手段
- 3 検索手段
- 4、94 順位づけ生成手段
- 5 検索結果表示手段
- 6 検索用索引展開手段
- 7 参照元付き文書表展開手段
- 8 ハイパーテキスト文書管理手段
- 9 ユーザ
- 10 文書管理手段
- 11 文書検索装置
- 20 関連キーワード入力窓
- 50 順位づけ結果表示窓
- 60～62 単語情報記入欄
- 70～72 文書情報記入欄
- 80～89 文書
- 800～806 アンカー文

【図1】



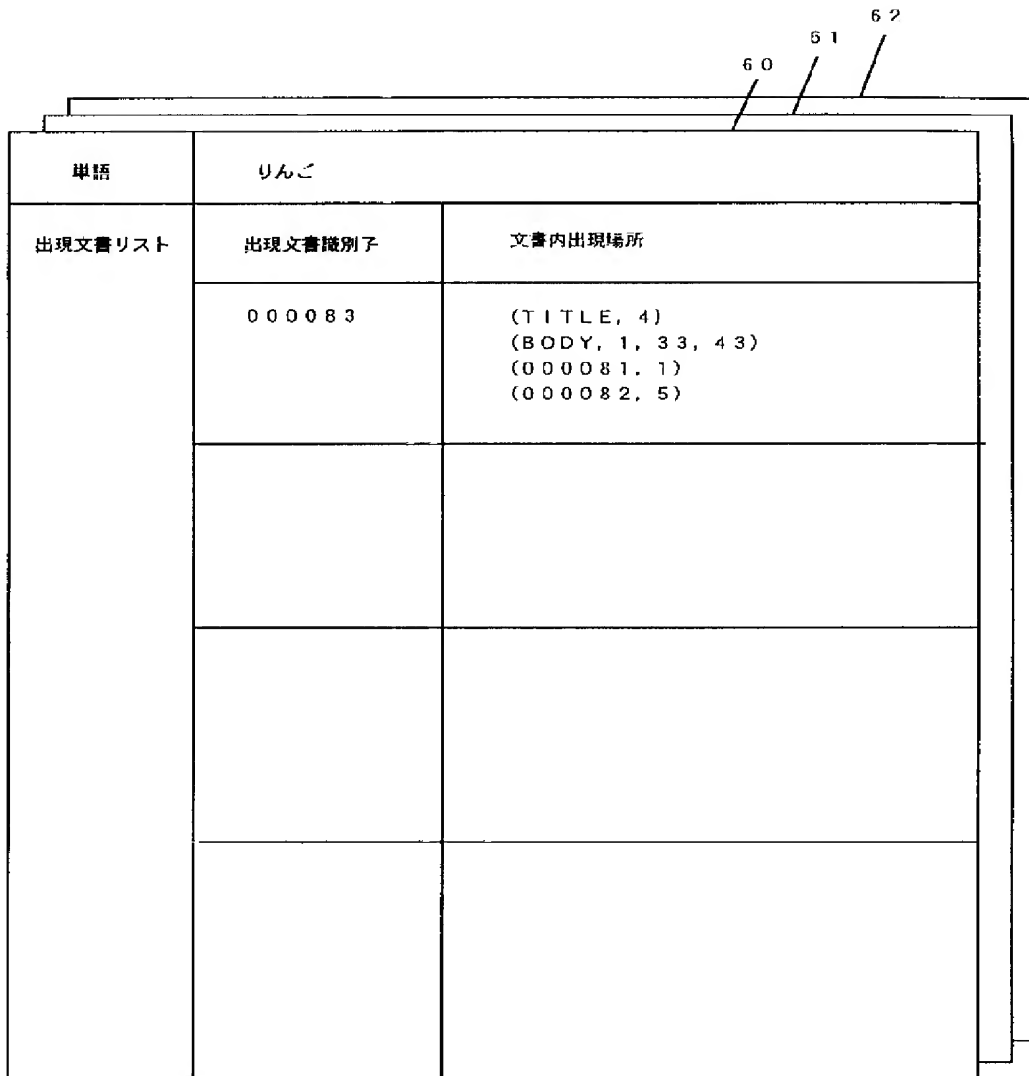
【図2】



【図3】

文書識別子	000083	
文書格納位置	www.mei.co.jp/home/tokusan/aomori/ringo.html	
タイトル	うらのりんご	
テキスト本文	りんごがたくさんとれました。 おいしそうですね。 農場の場所案内。 りんごの作り方。 宅配りんご、注文の仕方。	
参照元リスト	参照元文書識別子	アンカー文
	000081	りんご生産農家
	000082	おいしいりんごもあります。
単語リスト	りんご おいし 農場 場所 案内 作り方 宅配 注文 仕方 生産 農家	

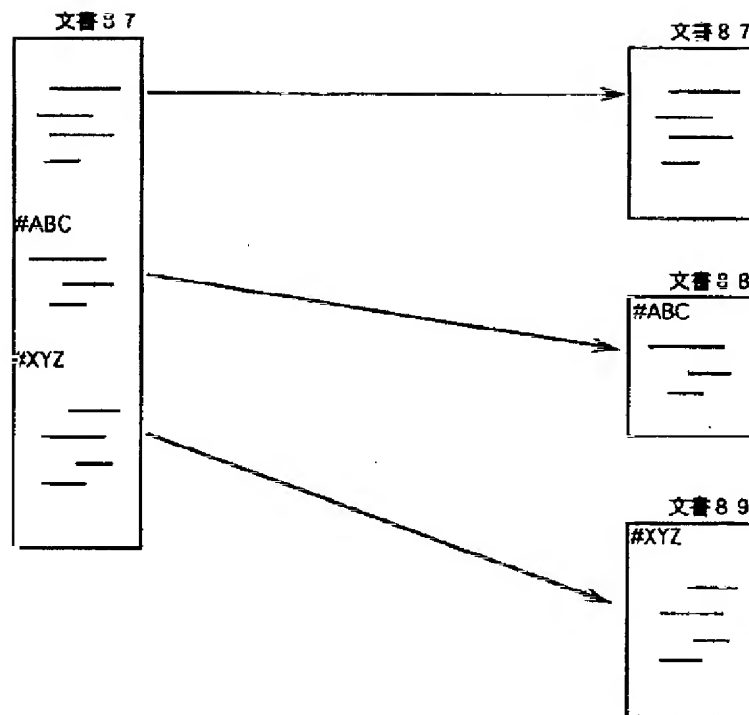
【図4】



The diagram illustrates a multi-sheet table structure. Three sheets are shown, with labels 60, 61, and 62 pointing to the top sheets. The table has three columns: '単語' (Word), 'りんご' (Apple), and '出現文書リスト' (Occurrence Document List). The 'りんご' column is further divided into '出現文書識別子' (Occurrence Document Identifier) and '文書内出現場所' (Occurrence Location in Document). The '出現文書リスト' column contains a list of documents, with the first entry being '000083'. The '出現場所' column for '000083' lists the following locations: (TITLE, 4), (BODY, 1, 33, 43), (000081, 1), and (000082, 5).

単語	りんご	
出現文書リスト	出現文書識別子	文書内出現場所
	000083	(TITLE, 4) (BODY, 1, 33, 43) (000081, 1) (000082, 5)

【図5】



【図8】

検索結果

検索キー [りんご]
86件中30件表示

・
・
・

[4]

文書 8 0 の見出し

文書 8 1 の見出し

文書 8 3 の見出し ☆

文書 8 5 の見出し ☆

文書 8 6 の見出し ☆

・
・
・

【図6】

検索結果

検索キー 「りんご」
86件中30件表示

・
・
・

[4]

文書80の見出し

文書81の見出し

文書83の見出し ☆

・
・
・

[18]

文書81の見出し

文書83の見出し

文書85の見出し ☆

[19]

文書81の見出し

文書83の見出し

文書86の見出し ☆

【図7】

検索結果

検索キー 「りんご」
86件中30件表示

・
・
・

[4]

文書80の見出し

文書81の見出し

文書83の見出し ☆

文書83の要約文

・
・
・

[18]

文書81の見出し

文書83の見出し

文書85の見出し ☆

文書85の要約文

[19]

文書81の見出し

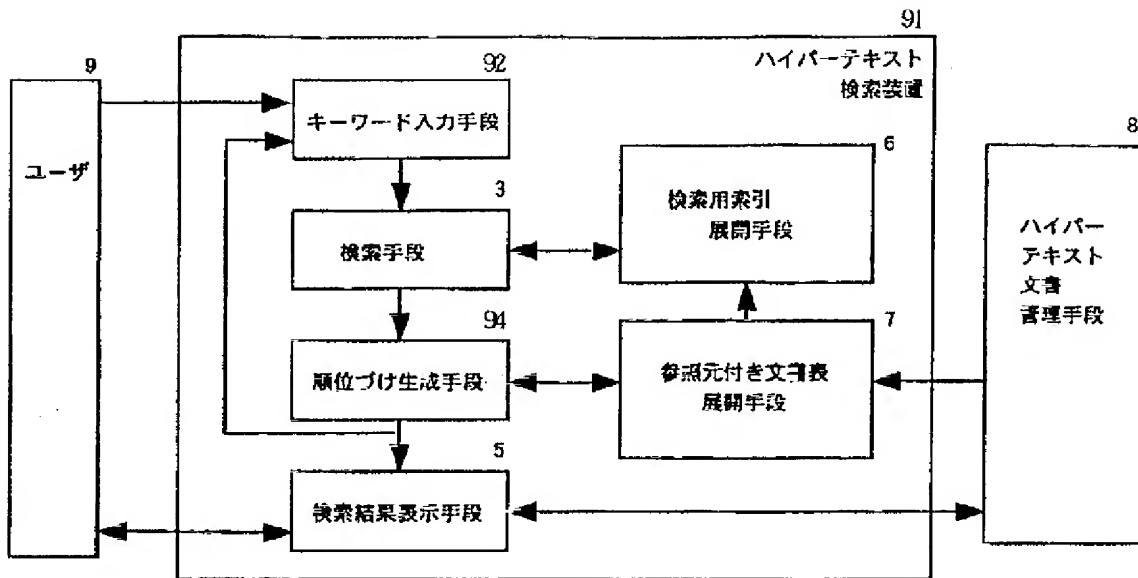
文書83の見出し

文書86の見出し ☆

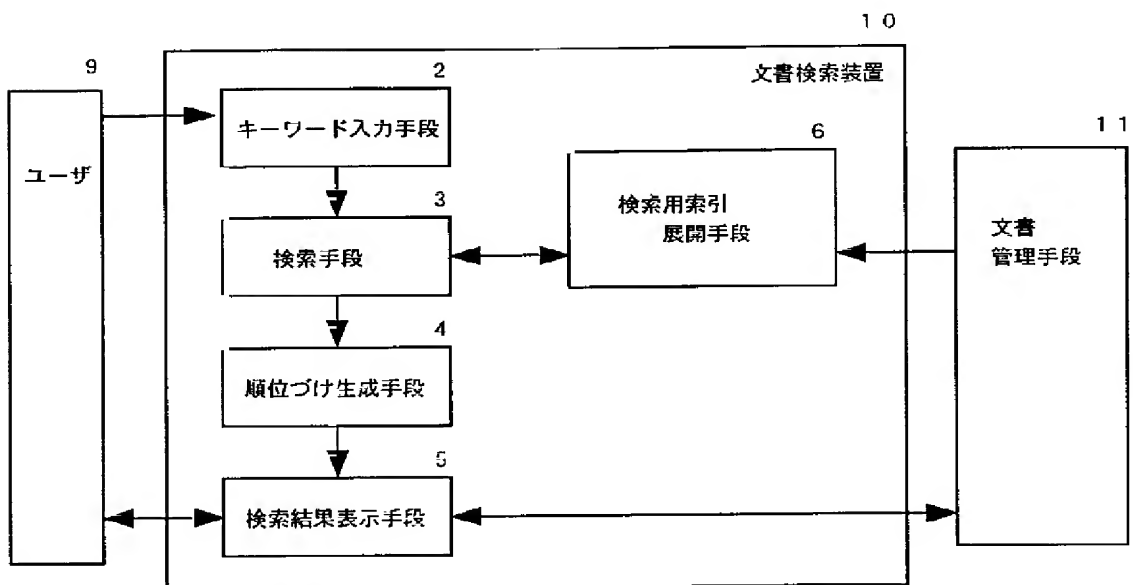
文書86の要約文

・
・
・

【図9】



【図11】



【図10】

検索結果

検索キー - [りんご]

86件中30件表示

...

[4] 文書80の見出し

文書81の見出し

文書83の見出し ☆

文書85の見出し ☆

文書86の見出し ☆

...

50

りんご

クリア

初期状態

再検索

信州

農家

特産品

青森

肥料

農園

祭り

ねぶた

長野

オリンピック

20

フロントページの続き

(51)Int.Cl. ⁶	識別記号	片内整理番号	FI	技術表示箇所
G06F 17/21			G06F 15/20	590E
			15/40	370A
			15/403	370Z

(72)発明者 野口 喜洋	(72)発明者 安川 秀樹
大阪府門真市大字門真1006番地 松下電器	大阪府門真市大字門真1006番地 松下電器
産業株式会社内	産業株式会社内